DOCUMENT RESUME

ED 195 581 TM 810 035

AUTHOR Balasubramonian, K.
TITLE Measurement, Evaluation and Accountability in Bilingual Education Programs.
PUB DATE 74
NOTE 28p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; *Bilingual Education; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; Measurement; *Program Evaluation; Summative Evaluation

ABSTRACT

A case for evaluation is developed and procedures for evaluating bilingual education programs are suggested as a guide for the novice evaluator. The following areas are discussed from the point of view of an evaluator: (1) how to plan for evaluation; (2) how to use needs assessment data; (3) the relationship between formative and summative evaluation; (4) the instrumentation process; (5) data collection; (6) data analyses; and (7) production of final reports. A brief section on the more recent consciousness for accountability points out how each of the participants in a bilingual education program can contribute towards accountability. A distinction between research and evaluation is made in order to alert the administrator on the most appropriate use of the information obtained from each. A bibliography, suggesting other sources of information, is appended. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195581

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Balasubramonian

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MEASUREMENT, EVALUATION AND ACCOUNTABILITY IN BILINGUAL EDUCATION PROGRAMS

K. Balasubramonian
1974

Introduction

The purpose of this article is to develop a case for evaluat. and to suggest procedures for evaluating bilingual education programs. Public education whether funded by the local education agency (LEA), the state education agency (SEA) or a federal agency, uses public monies with a promise that the stated objectives will be met for every participating child. Each and every participant in the educational program, therefore, the LEA, the teacher, the student and the parent, has a defined or potentially defined responsibility in achieving the program objectives. The educator who designs or initiates a particular program, however, is further responsible for providing evidence as to the achievement of the stated goals. This is the basic idea of accountability. Accountability may be towards the public, towards a professional group such as the Office of Education, or towards the educator himself and his professional desire to determine the effectiveness of his program.

Whatever type of accountability may be desired, the means to obtain evidence as to the achievement of the objectives are measurement and evaluation. While measurement is concerned with obtaining data on student status and growth, and information on program characteristics, evaluation is concerned with assessing the merits of student performance and program characteristics. Evaluation, though different from research, has been considered one aspect of research; namely, the uncovering of durable relationships between pupil achievement, evaluation and

TM 810 035

accountability, then, are three interrelated concepts which play a definite role in establishing the merits of the specific educational program and in identifying instructional strategies which are most conducive to attaining the program objectives.

The following areas will be discussed from the point of view of an evaluator: how to plan for evaluation, how to use needs assessment data, the relationship between formative and summative evaluation, the instrumentation process, data collection, data analyses and production of final reports. A brief section on the more recent consciousness for accountability points out how each of the participants in a bilingual education program can contribute towards accountability. Finally, the distinction between research and evaluation is made in order to alert the administrator on the most appropriate use of the information obtained from each. To conclude, this article is written as a guide for the novice evaluator or program coordinator who has been assigned the task of an evaluator. The bibliography suggests other sources of inofrmation avaiable.

Planning for Evaluation

The goal of educational evaluation, in general, is to arrive at scientifically valid conclusions that will lead to the improvement of student learning. Since this can be a complex process, careful planning should determine the parameters in view of the specific project. The following activities and sample situations whenever possible will illustrate initial issues to be considered.

Decide on specific purposes of evaluation. Posing the question, "How will the information obtained through evaluation be used?", various answers may be: to disseminate information, to assist administrative officials make appropriate decisions, or to modify the program. While any or all three may be identifiable purposes, the focus should be decided in the initial stages.

For example, the role of evaluation for the Title VII programs as indicated in the guidelines for the past few years has been primarily informative, secondarily for the benefit of decision-makers, and least of all for the benefit of those interested in modifying programs at the local level.

The Title VII evaluation reports provide information as to the extent to which individual program objectives under specified components have been achieved. In this sense, the role of this evaluation is informative. The mandatory (until 1972-73) internal audit is a cross-check on the reliability and validity of the evaluation. Based on the information reported, decisions were made to increase or decrease funding for particulat projects or develop new sites, and to revise program guidelines. At the local level, school districts may have used the information as a basis for modifying individual programs; however, this was not the intended role of the general Title VII evaluation. It may also be noted that, insofar as this evaluation was aimed solely at providing information and in establishing permanent relationships, it was performing a research function.

Whatever the focus of evaluation, then, the decision should be made according to pre-established guidelines and consequent meetings with the funding agency. At this time, the choice of variables, designs and processes of evaluating may be clarified. For example, if program modification is one of the intended roles of evaluation, the means for assessing the effectiveness of the instructional process should be included.

Identify variables of concern and project objectives related to them. If the purpose of evaluation is to assist in the decision-making process or to modify a program, all the variables related to the program should be included in a comprehensive evaluation design. This, however, will make the evaluation process a complicated, difficult task; and, as a consequence, tend to increase the chances of drawing less reliable data. Specifying too few variables, will also yield inadequate data. The scope of evaluation therefore should be determined according to the stipulated guidelines and on the funds, personnel and facilities available for this purpose.

The best procedure for selecting important variables would be to review the project proposal and meet with all personnel involved in planning and implementing the program. Appropriate choices can then be drawn from a general listing of questions to be answered. In addition, characteristics of the project participants and personnel should be included as related variables; information on these can be secured from school records, school district and national surveys, and the local needs assessment that may have been done for the project proposal.

The guidelines for the bilingual education program in the United States have traditionally identified five major components to be evaluated: instruction, staff development, materials development, community involvement and management. In the past two years, Title VII evaluation guidelines have been limited to assessment of the instructional product. The State of Illinois requirements, on the other hand, have placed more emphasis on instrumentation analysis, instructional product evaluation through design modification, and instructional process evaluation through external audit.

To summarize, the following chart presents the process.

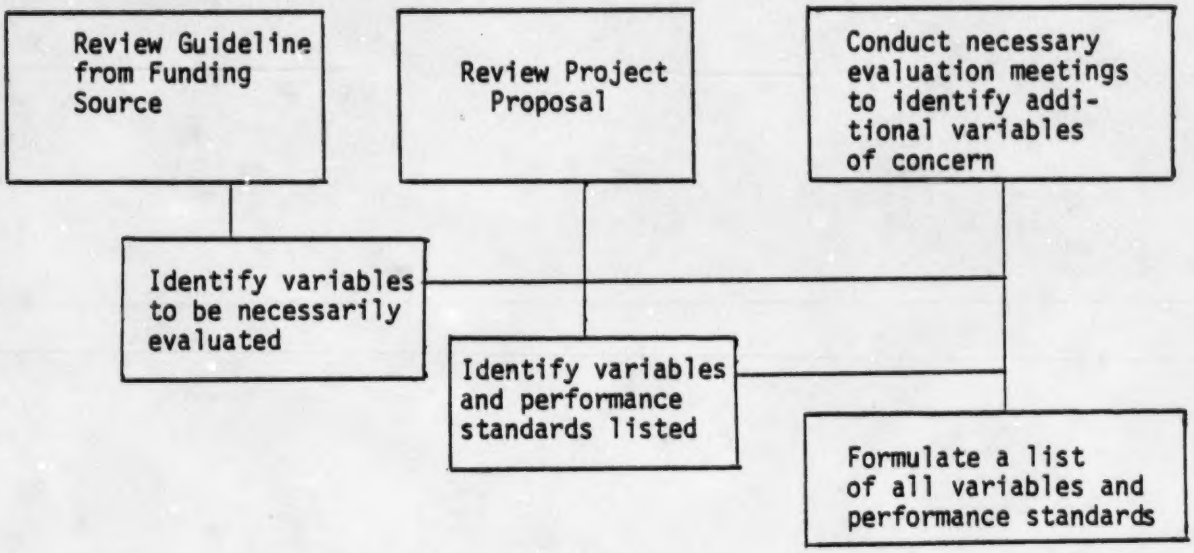

Review Guideline from Funding Source
Review Project Proposal
Conduct necessary evaluation meetings to identify additional variables of concern
Identify variables to be necessarily evaluated
Identify variables and performance standards listed
Formulate a list of all variables and performance standards

Figure I: Identification of Variables and Selection Process

Specify evaluation design, including plans for instrumentation. Once the purpose of evaluation and the variables to be included are established, the ground is set for drawing an evaluation design and selecting the evaluation instruments to be used. If the purpose is to disseminate information or to modify a program, the evaluation design is simple. Pre-post and process measures may be collected; a comparison with stated standards of performance may be done; and necessary modification of the process or standards may be made.

If the focus is on accountability, the evaluation will be more meaningful if a pre-post measure on an experimental-control group design is used to assess pupil achievement. The problem of establishing control or even comparison groups, especially according to true experimental design standards, is very difficult and sometimes impossible.

Recently, the Bilingual Education Section of the Illinois Office of the Superintendent of Public Instruction (OSPI) has developed an innovative design modification technique to overcome this difficulty. Participating school districts are given the choice of the following six designs for comparison ranging from rigorously true experimental designs to quasi-experimental designs,

Design I. The ideal experimental design would be one that permits random assignment of closely-matched subjects to treatment-control groups. This model has been proposed in theory only; since pupils are assigned to the program because of optimum needs, comparable groups cannot generally be established.

Design II. This true experimental design is classified as "random within stratum." Pupils in greatest need for bilingual education are automatically placed in the program; their progress is assessed but not compared. Pupils in a "less needy" category are randomly assigned to experimental and comparison groups; only data from students in this category is used for comparison.

Design III. Comparison groups in this "parellel groups" design are drawn from classes or schools having no bilingual program, but matching program participants in a number of socio-economic variables.

Design IV. This "regression-discontinuity" type design consists in comparing scoring trends of bilingual program pupils in specific grade levels (i.e., P1, P2, and P3) with that of non-program students in adjacent grade levels (i. e., 4th and 5th). The design is based on the principle that the relation between two variables may be statistically predicted. The term "regression" is used in its statistical meaning, not in its psychological connotation.

Design V. This "grade-cohort" design is best suited for programs in the first year of operation. Since test data in adjacent levels overlap without systematic bias, pretest scores for one grade level (i.e., 5th) may be compared with postest scores of a lower grade level (i.e., 4th).

Design VI. In this "stratified student population" design, different populations are compared for their contrastive interest. For example, test data for Anglo students in the bilingual program may be compared with that of Anglos not in the program.

The projects choosing a particular design must be able to meet the requirements. For example, the grade cohort design requires that a program be in its first year of operation. Since projects may differ in program structure, specific objectives, facilities and amount of funding received, the OSPI multiple-design allows for individual differences within a coordinated evaluation. Accurate record keeping is essential to the successful implementation of this model.

The instrumentation aspect of the evaluation design is a delicate process and deserves careful planning, especially in the case of bilingual education programs. In a later section this process will be treated in greater detail.

Establishing a time line. The last but equally important step in planning for evaluation is to establish a projected time line and allocate responsibilities. The procedure includes identifying available personnel and each evaluation activity is to be completed and finally assign responsibilities to appropriate personnel. The dates should be realistic estimates of time required for the completion of each task.

To conclude, all four activities described in this section constitute the planning stage of the program evaluation. Its success depends on the efficiency of its management. The reader is referred to the Program Evaluation and Review Technique (PERT) for further suggestions on efficient planning and management.

Needs Assessment: A Prerequisite

In the planning for evaluation it is assumed that the program structure and participants have been selected on the basis of a systematic needs assessment. The process is mentioned here, however, because it is generally the role of the evaluator to show evidence of the need for the project and to validate each participant's membership on the program. In many cases, this dual role of the evaluator--needs assessor and program evaluator--has caused misuse of the needs assessment data. For example, in the interest of conserving valuable time and effort, a formal pretest administration may be averted by using data from a selection test. This procedure is generally not recommended unless appropriate steps are taken to correct possible statistical regression effects.

Formative Evaluation: for the Practitioner

Information from evaluation is mainly for two types of interest groups, the practioner (teachers and curriculum developers) and the administrator. The practitioner in general is interested in the evaluation of the instructional product and process. He seeks answers to questions such as:

- How well have the students performed in this area? What are their weaknesses and where should more effort be exerted?
- What are the identifiable weaknesses in the curriculum and/or in the instructional method used? How can these be modified to suit the students on class?

The concept of formative evaluation was developed to provide information that would clarify these issues.

Stated another way, formative evaluation seeks to improve student learning. Its major function is to trace student growth by determing the degree of mastery achieved in a learning task and pinpointing the part of the task not mastered. The purpose is not to help the teacher grade the learner but rather to keep both focused on moving toward mastery. A second function is to provide a basis for modification of curriculum materials and instructional methods.

The concept of "mastery learning" deserves further clarification. Philosophically, it is based on the humanist ideal that every child, if given sufficient time to do so, is capable of mastering a given task. Two implications here are that 1) no learning is complete until mastery is achieved, and 2) a person's aptitude may be determined by the time he requires to achieve mastery.

Instruction for mastery learning may be extreme individualized instruction; however, a basic difference is that the latter is often based on the measured ability of the participants, whereas the former was based on the time required to learn a given task.

Whatever approach is used for the instructional process (individualized, mastery learning, remedial or regular), formative evaluation is of utmost importance for the practitioner. The process may be illustrated as follows:

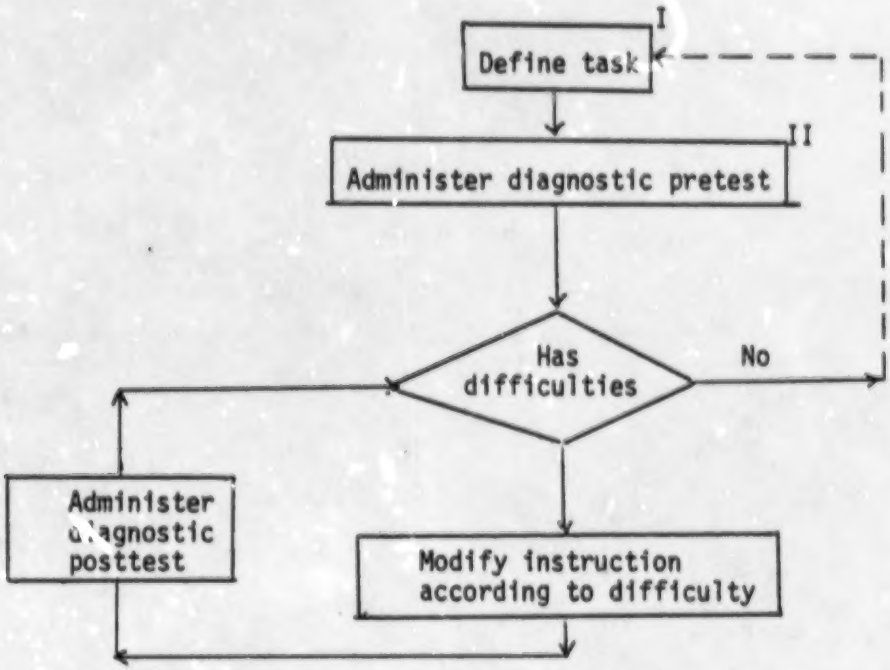


Figure 2: Formative Evaluation for Instructional Product

The first step in formative evaluation (and in the instructional process as well) is to define the task; that is, formulate specific behavioral objectives with appropriate standards of performance. To illustrate, a specific objective for a lesson in map reading may be

> Given a map of Illinois, and given the row and column (e.g., A10 or H22) of a number of places, the student will locate each place in the map correctly 80% of the time.

The second step involves diagnostic pretesting. Such a test will have a number of items to measure some of the specific objectives. As in the example cited, the test might include five or six specific objectives, each being measured by five or ten items. If a pupil passes 80% of the items under each objective, he may be considered to have mastered the task of map reading. It is possible that in the first diagnostic posttesting (where a parallel form of the protest is used), the pupil may show mastery of three of the five objectives. This indicates

diagnostic posttesting (where a parallel form of the pretest is used), the pupil may show mastery of three of the five objectives. This indicates further instruction is needed for the remaining two objectives. A diagnostic posttest on these objectives may then be given. This procedure, called "criterion reference testing", is often used in formative evaluation. The nature and charactieristics of criterion reference measures are listed in the section on instrumentation.

Summative Evaluation: for the Administrator

While the teachers are interested in applying the results of evaluation to classroom processes, the administrator (principals, superintendents, HEW, etc.) is interested in answering questions such as:

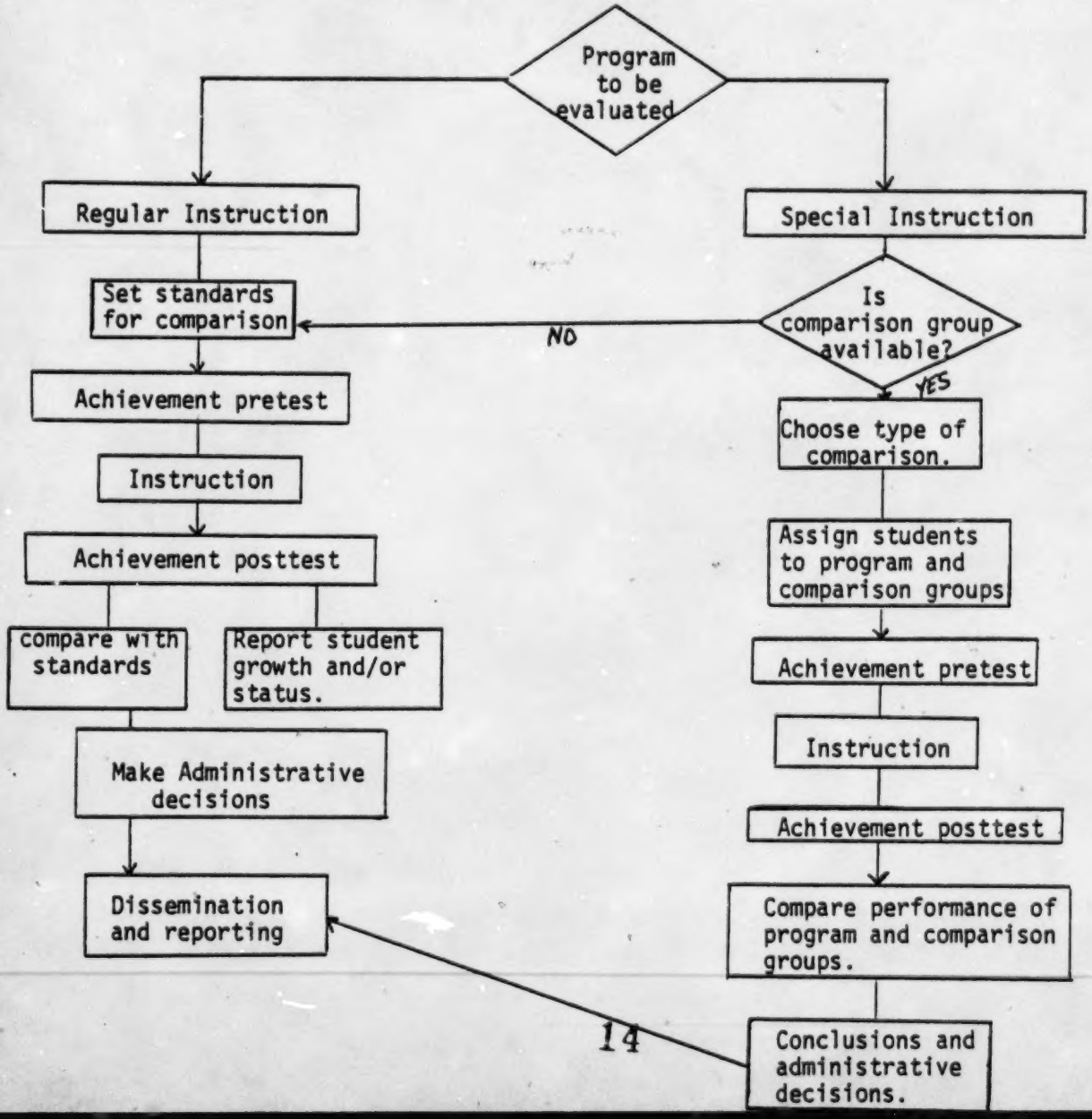
- What is the overall effect of the program on the students?
- What additional benefits do the students receive for the additional funds spent in their education?
- What are the strengths of the program and how can it be improved?

To answer such broad, general questions requires an evaluation that is based on true experimental designs such as pre-post comparison between closely-matched groups that have been randomly assigned to treatment and control situations. Or, at least a comparable quasi-experimental design along with process evaluations are needed. This process is known as "summative evaluation" representing a total view of the nature and overall effect of the instructional program. The purposes are to evaluate

student growth over a relatively long period of time, to report student progress, and make recommendations that may contribute to a sound policy-making process. This type of evaluation is generally used in regular instructional programs or special educational programs.

The procedure for a summative evaluation are presented in the following diagram (Fig. 3). One is suited for regular instructional programs; the other for special education programs. The two differ only in the initial step and the kind of statistical analysis used.

Figure 3: Summative Evaluation for Instructional Product



From the description of the two types of evaluation, it can be noted that both are necessary components of a comprehensive evaluation; hence, both should be used in a complementary fashion. Further, it should be noted that the two types together evaluate only the instructional product component of the total program. Other components (instructional process, curriculum and staff development, community involvement and management) may be evaluated either for descriptive purposes or to establish comparative and/or equivalent influences on program and comparison groups.

When change-inducing activities (i.e., staff development, community involvement), appropriate analyses should be performed? Sometimes it may be necessary to compare changes in the affective area which may have been due to staff and community activities. The choice of components and variables should be the joint decision of the administration and evaluation staff at the initial conference.

Instrumentation: A Vital Problem

This process generally requires intensive planning and concentrated effort. It calls for the selection, adaptation and/or development of appropriate instruments to measure each variable in question. Oftentimes, special computerized forms are necessary to collect and bank program statistics and student background information. Student rating scales, parent and teacher attitude scales, inservice needs questionnaires, etc. may need to be developed. It is strongly recommended that such instruments be constructed by or with assistance of professionals in the field.

The major difficulty under instrumentation is generally the selection of standardized achievement tests, especially in the case of bilingual education programs. The participants in these programs are selected, cultural-ethnic group children who have minimal (if any at all) skills in the English language and are culturally different from the general U.S. population. Further, a considerable fraction of these children are not even familiar with the test formats and procedures that are common in the U.S. educational system. While there are valid, reliable achievement tests available, these are not appropriate to use with bilingual student populations.

However, standardized tests are necessary for making nomative judgements and meaningful comparisons. A compromise situation, then, could be a re-analysis of test characteristics, a re-norming process, and interpretation of the data in terms of both, the publisher's norms and the norms for the reference group. Such a procedure of instrumentation analysis has been instituted an integral part of evaluation for Illinois bilingual programs.

Criterion referenced (CR) tests for formative evaluation may be standardized on a set of specific objectivism common tests developed by school districts or even other teacher-made CR tests. Care should be taken that such tests be geared to measure the specific objectives to be achieved by the child. Usual test characteristics (i.e., reliability, validity, etc.) must also be considered when selecting a test. Unless the objectives are standard objectives, the use of published criterion referenced tests is not advisable. Teacher-made CR tests are more useful in such cases.

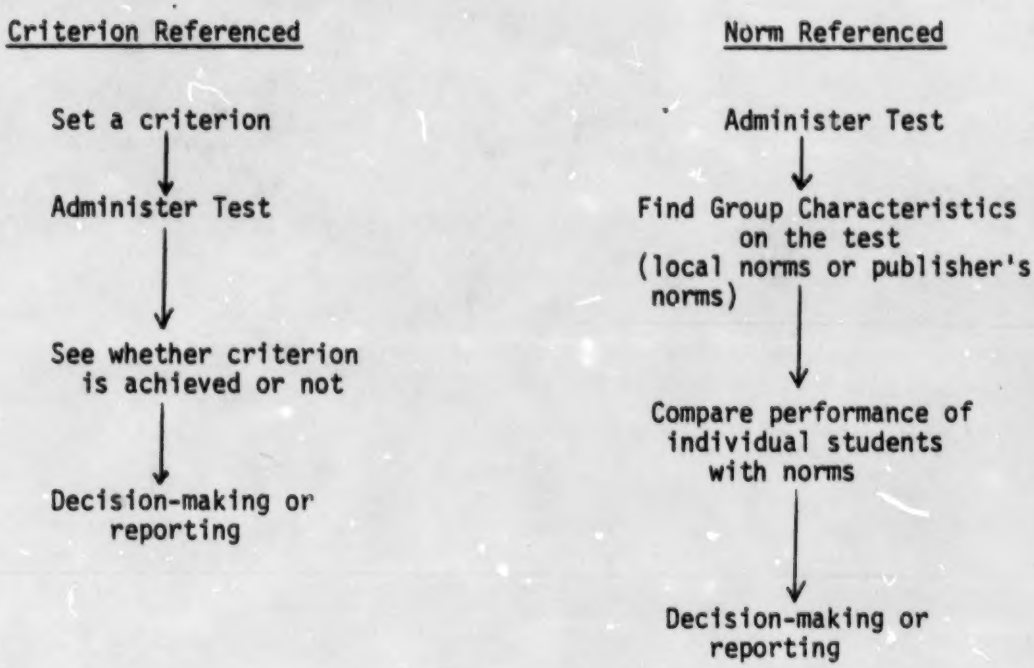
Figure 4

COMPARISON OF SOME TYPES OF TESTS

Type / Characteristic	Placement	Diagnostic	Criterion Referenced	Unit	Achievement
Scope	broad	limited (usually)	no restriction	limited	broad
Number of items	large	limited	no restriction	limited	large
Purpose	Assess general level of performance.	Assess student difficulties in specific instructional areas.	Evaluate student performance with reference to a present criterion and not in reference to group norms.	Assess achievement in specific instructional units.	Assess general level of performance.
Interchangability	May be used as achievement test for adjacent levels.	None	None	A collection of sequential unit tests may be used as placement tests.	May be used as placement tests for adjacent levels

Figure 5

USING CRITERION REFERENCED AND NORM REFERENCED TESTS


Criterion Referenced
Set a criterion
Administer Test
See whether criterion is achieved or not
Decision-making or reporting
Norm Referenced
Administer Test
Find Group Characteristics on the test (local norms or publisher's norms)
Compare performance of individual students with norms
Decision-making or reporting

Information and reviews on standardized achievement and criterion referenced tests are available in the Seventh Mental Measurements Yearbook and the Center for the Study of Evaluation Test Evaluations.

Criterion referenced tests can function as diagnostic tests; but standardized diagnostic tests are also available. Figure 4 on the following page presents a comparison of some type of tests in terms of their content scope, number of items, purpose and interchangeability.

It should be noted also that norm referenced and criterion referenced tests may be interchangeable, provided appropriate data is available. It is the role in which the test is used that determines its basic characteristic. Figure 5 presents a comparison between the two types and how each is used.

Although it is a difficult and time consuming task to develop a test, teacher-made tests have definite advantages, especially for formative evaluation because they have reference to immediate classroom objectives and curriculum, and because the teacher can determine the length to suit local needs. However, teacher-made tests are not generally suited for comparison between groups because they lack refineness and test characteristics cannot be determined, and because of probable bias to the non-reference group due to difference in program objectices and curricula.

To conclude, teachers--especially those involved in bilingual education--are encouraged to work comparatively in developing teacher-made tests. A group of third-grade teachers from various bilingual centers may get together to write items for a social-cultural studies test. Probably 75% if not all will have had a background course in tests and measurements from which to draw. Other sources of technical assistance may be sought through local school or community organizations. Educational Testing Service measurement kits and item pools in the various content areas and grade levels may also be readily available.

Data Collection: A Vital Step

Test administration and data collection are two steps in the evaluation process which need close supervision. Careless or "helpful" proctors, non-cooperative respondents, non-acceptable testing conditions, vague instructions, are some of the factors that can invalidate the data collection process and thus the entire evaluation. Three areas to consider, then, are preparation, administration and processing.

Preparation. A prerequisite for administration is that both tester and testee be properly prepared. This means that proctor must familiarize himself with the testing materials and specific instructions a day or two prior to actual testing. When a battery of tests is being used, the compilation of specific instructions into a manual may save time and effort. A set of general instructions for the test administration and a testing schedule blank could be included. Evaluation for the bilingual programs in the Chigaco and down state Illinois schools have found this practice most helpful. The evaluation plan also includes a general

orientation session--attended by all testing proctors--at which the overall evaluation is explained, instruction manuals are distributed, and sample tests are presented.

The testees may be prepared by a brief statement on the purpose (real or disguised) of the testing program and the need for their cooperation. Testees who may not be familiar with the test format should be given a practice test with two or three sample items in each category prior to actual testing.They should also be familiarized with the computer-type answer sheets and information blanks.

Administration. Test administration should be done in the most ideal or mear ideal conditions. For individual testing secluded rooms are essential. In group testing, the number of individuals that should be tested at one sitting depends on the capacity of the room and number of proctors. Groups of five for preschool and kindergarten children are advisable. Groups of 20 at the primary and intermediate levels, groups of 30 beyond grade six are suggested.

It is important that the testee understands the task that he is required to do, whether it be filling out identification data or answering the subtests. This is the proctor's responsibility. It is advisable that, for young children, identification blanks be completed prior to testing. Time limits should also be announced at the proper times. Most important to encourage students "do their best" is to generate a relaxed but firm atmosphere.

One means of supervising proper test administration is to request that proctors complete, after each testing session, a "test inventory" which describes the characteristics of the testing situation. Another method is through audit procedures whci may be part of a general audit of the total evaluation process. An independent audit is especially useful for summative evaluation.

Processing and Scoring. Each completed test must have proper identification on the student, the program and the testing session. Unambiguous test scoring keys must be provided. In the case os essay and short answer tests, appropriate scoring guidelines should be available and efforts to improve the reliability of the scoring must be made. Machine-scorable tests save time and increase the reliability of scoring. If scoring facilities are not avialable, the task may be contracted with independent agencies. In this sense, appropriate budget allocations are to be made.

To conclude this section, when data are to be collected from parents and community, the suggested procedure is through personal interviews. When funds and personnel are lacking, parent-community meetings in the school may be used for interviewing. The least recommended method is sending questionnaires through the mail. A combination of these may certainly be used.

Data Analysis

Once all tests are scored and other data is coded and organized, data analysis may be done for appropriate interpretations. This stage is more crucial for summative evaluations. In formative evaluation the only judgement that is to be made is to see whether the testee has met

the set criterion of mastery for each of the specific objectives. No statistical analysis of data is necessary.

Summative evaluation calls for group data and comparative judgements. The type of statistical analysis performed will depend on the hypothesis to be tested and the type of comparisons available. The most commonly-used data treatment procedure is converting the raw score into grade equivalent and comparing the average grade equivalent gain in the program and comparison groups. When a number of grade-equivalent-months gain expected through the program is specified, and a comparison group is not available, the simple analysis shown in Figure 6 may be done. Many other statistically superior but more complex data analysis procedures are available; if this is desired, technical expertise may be contracted.

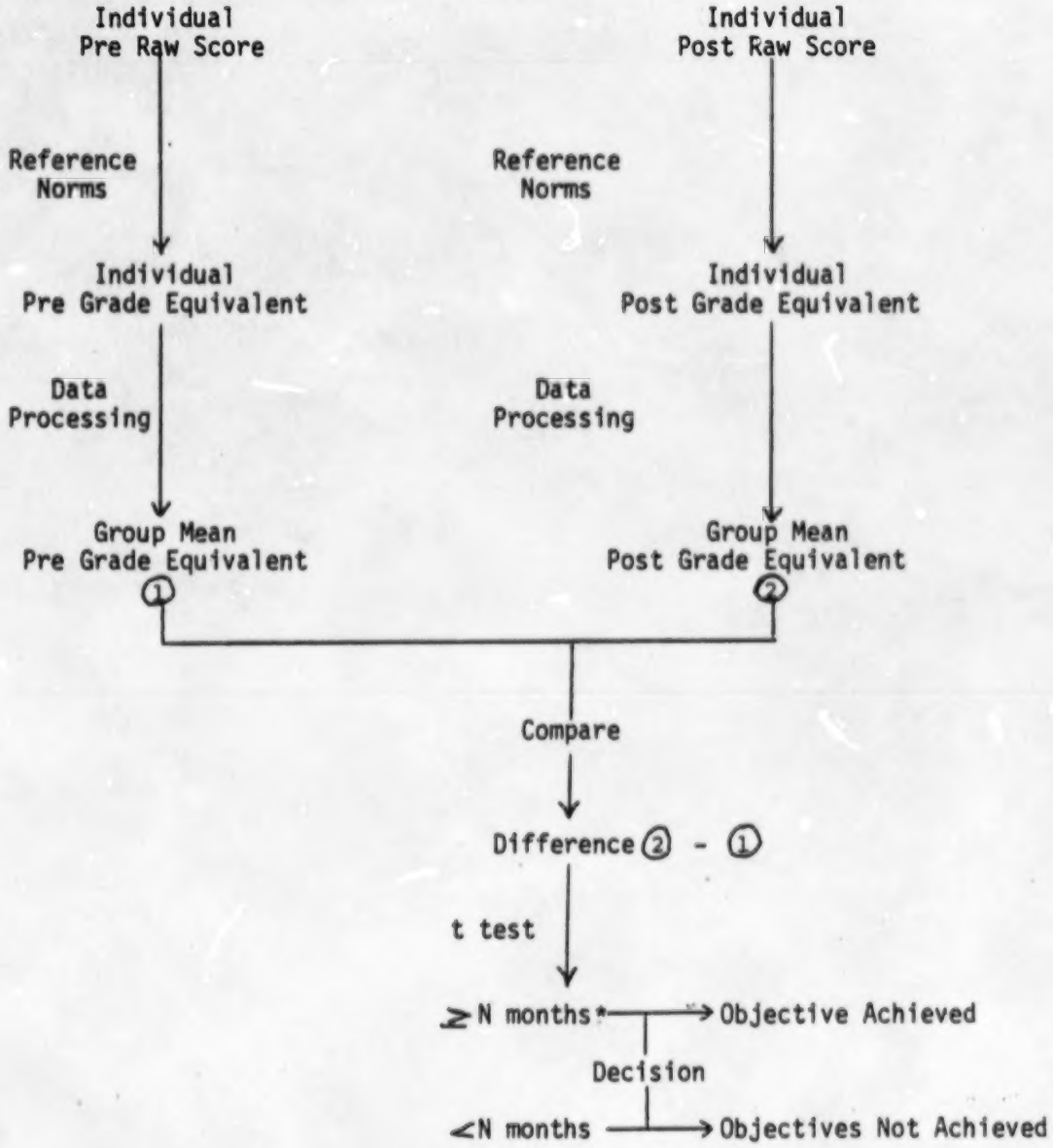
Final Reports: Not the Last Step

Final evaluation reports are essential and avluable dissemination instruments. These reports generally contain a brief history and description of the program being evaluated, its objectives, the evaluation design and data analysis procedures used, the findings, and recommendations for program modification. Depending on the scope of the specific evaluation, the recommendations may be broad or specific. When the program extends over a long period of time, it is advisable to have one or two interim evaluations reports.

As the heading suggests, the presentation of a final evaluation report should not be the last step in the management of the program. The recommendations contained therein are the result of concentrated effort from all those involved in the program from its imception and

Figure 6

PROCEDURE OF ANALYSIS



*N = number of months between Pre-and Posttest

implementation theough its evaluation; hence, they should not be glanced over quickly. The recent practice of on-site evaluation of processes and immediate feedback is quite valuable, as teachers tend to implement the oral feedback (followed by written reports) much more frequently. It is important, then, that any recommendations made should not be considered a threat to the practitioner or a challenge of the practice, but an aid in identifying aspects which may have been overlooked.

Accountability: Everyone's Concern

In the introduction, the case for evaluation was developed through the concept of accountability. However defined, accountability has meant "assigning responsibilities." In this sense, much controversy has arisen as to who should be responsible for what. A more flexible frame of reference might be to redefine the concept in terms of "a willingness to accept responsibility." Seen this way, most if not all advocators of a new program would probably subscribe to it. On-site feedback, interim and final evaluation reports are just sides to determine how far these responsibilities have been accomplished.

Conclusion

The purpose of this article was to give a broad overview of an evaluation model for bilingual education programs. An effort was made to keep technical information to a minimum. However, the reader is provided with a bibliography for further in-depth study of each section. Technical assistance may be necessary, especially when there are research interests in view. For example, a research question might be "How different are the various bilingual education models in terms of their effectiveness?" Questions like this require appropriate control of

variables such as comparability of schools, teachers, instructional processes, among other things. If funds and facilities are available, one or two research projects may be performed along with evaluation. However, it should be noted that research is not the purpose of evaluation, and that the two are distinct processes in their own right.

Balasubramonian is coordinator of field services at the Bilingual Education Service Center, Arlington Heights, Illinois.

BIBLIOGRAPHY

Avis, J. L. and W. M. Essey. Educational Evaluation: Official Proceedings of a Conference. State Superintendent of Public Instruction, Columbus, Ohio, 1969.

Balasubraminian, K., H. N. Seelye and R. Weffer. "Do Bilingual Education Programs Inhibit English Language Achievement?: A Report on an Illinois Experiment," in process of publication in the Modern Language Journal.

Bloom, B.S., J. I. Hastings and G. F. Nadaus. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw Hill, 1971.

Buros, O. K., ed. The Seventh Mental Measurements Yearbook, Vols. I and II. Highland Park, New Jersey: The Gryphon Press, 1972.

Campbell, D. T. and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1966.

A Comprehensive Evaluation Design for ESEA Title VII and Illinois Funded Bilingual Education Programs in the Chicago Public Schools, FY 1974. Board of Education of the City of Chicago.

Cook, D. L. Program Evaluation and Review Technique: Applications in Education. U. S. Department of Health, Education and Welfare, Office of Education, Washington, 1966.

Edwards, A. Statistical Analysis. New York: Holt, Reinhart and Winston, 1965.

ESEA Title I School Year 1972 Needs Assessment Survey. Board of Education of the City of Chicago.

ESEA Title VII Final Evaluation Report for Bilingual Education Programs in the Chicago Public Schools, FY 1972. Board of Education of the City of Chicago.

Johnson, O. G. amd J. W. Bommarito. Tests and Measurements in Child Development: A Handbook. Jossey-Bass, 1971.

Kaufman, R. A. amd J. R. Harsh. Determining Educational Needs: An Overview. ERIC No. ED 038 551, 1969.

Mager, R. F. Preparing Instructional Objectives. Belmont, Cal.: Fearon Publisher, 1962.

Mehrens, W. and R. L. Ebel. Principles of Educational and Psychological Measurement: A Book of Selected Readings. Chicago: Rand McNally, 1967,

Mehrens, W. and Lehman. Standardized Tests.

Multilingual Assessment Project - Test Reviews. Stockton, Cal.

Seelye, H. N. and K. Balasubramonian. "Accountability in Educational Reform Programs through Instrumentation Analyses and Design Variation: Evaluating Cognitive Growth in Illinois Bilingual Programs." EDRC

Stake, R. E. "The Countenance of Educational Evaluation," Teachers College Record, 68(7) April, 1967.

State Educational Assessment Programs, Educational Testing Service. Princeton, N.J., Berkeley, Cal., Evanston, Ill., 1971.

Stenner, A. J. An Overview of Information Based Evaluation: A Design Procedure. Arlington, Virginia: IBEX, 1972.

State of Illinois Final Evaluation Report for Bilingual Education Programs in the Chicago Public Schools, FY 1972. Board Of Education of the City of Chicago.

Test and Measurements Kit. Educational Testing Service. Princeton, N,J,m 1973.

Thorndike, R. L., ed. Educational Measurement, Second Edition Washington, D.C.: American Council on Education, 1971.

Tyler, R. M., ed. Educational Evaluation: New Roles, New Means. Sixty-Eight Yearbook of the National Society for the Study of Education, Part II. Chicago, Ill: University of Chicago Press, 1969.

Webb, E. J. et al. Unobtrusvie Measures: Non-Directive Research in the Social Sciences. Chicago: Rand McNally, 1966.

Wrightstone, J. W., J. P. Hogan and M. M. Abbott. Accountability in Education and Associated Measurement Problems. Test Service Notebook 33. New York, Harcourt Brace and Jovanovich.